NON-INVASIVE CLASSIFICATION OF CORTICAL ACTIVITIES FOR BRAIN COMPUTER INTERFACE: A VARIABLE SELECTION APPROACH

Michel Besserve, Jacques Martinerie and Line Garnero

Laboratoire Neurosciences Cognitive et Imagerie Cérébrale,
CNRS UPR 640 LENA & UPMC Univ Paris 06, Paris, FRANCE

ABSTRACT

We propose to carry out a classification method for electro-encephalographic signals (EEG), using the activities of cortical sources estimated with an EEG inverse problem. To overcome the difficulties caused by the high number of sources (approximately 10000), we use a multivariate variable selection algorithm: the zero norm Support Vector Machine (L0-SVM). This technique allows to extract a small subset of sources, which are the most useful to allow for the discrimination of the mental states. The whole approach is applied to an asynchronous Brain Computer Interface (BCI) experiment from our lab. It outperforms a method based on the direct measurement of EEG electrodes' activities.

Index Terms— EEG, Inverse problem, Brain Computer Interface, Support Vector Machine

1. INTRODUCTION

Brain computer interfacing (BCI) is a challenging application aiming at providing a communication channel for completely paralyzed patients through the real time analysis of their cerebral functional imaging data during specific mental tasks. This analysis is usually performed in two steps: 1) quantification of cerebral imaging signals to extract informative characteristics of cerebral activity, 2) classification of the quantified data to estimate the mental task being performed. The most promising non-invasive BCI devices are based on electro-encephalographic (EEG) signals because of the high temporal resolution and portability of the recording system. Nevertheless, the information transfer is limited by a low spatial resolution, as the signal on each EEG sensor can originate from several brain regions. To overcome that, source reconstruction methods enable to estimate the electrical activity of the populations of cortical neurons from EEG measurements. According to the literature on anatomo-functional organization of the cortex, this can lead to a better dissociation of the activities of many task-specific neural networks at this "source level" compared to the "sensor level". As estimating cortical source activity requires to solve a largely under determined inverse problem, many solutions have been proposed in the literature. Among them, distributed linear inverse solutions such as minimum norm estimate (MNE) [1] are advantageous because computationally tractable in the real time context of BCI. These methods enable to estimate simultaneously the amplitude of many current elements uniformly distributed over the cortex using a linear transform of the instantaneous EEG signals. One shortcoming of this approach is the large amount of equivalent dipoles (usually 10000) in the model, generating a high number of variables in the quantification step of the BCI device. This leads to an increase of computational time compared to a usual sensor level analysis and possibly lowers the performance of statistical methods such as classification algorithms. Moreover, these inverse solutions are known to generate spatially correlated cortical activities. Thus selecting a subset of informative quantification variables can be sufficient to estimate correctly the mental state of the subject. To reduce the number of quantification variables, variable selection techniques can be used to select the most useful subset of variables for the classification algorithm. Two approaches can be distinguished: univarate and multivariate variable selection [2]. In univariate variable selection, the discriminant power of each variable is evaluated separately, using for example the Fisher score, and is used to rank the variables. The best ones are then selected for the classifier. At the opposite, multivariate variable selection takes into account the variable space as a whole to select the best subset of variables. The second approach is particularly well suited in case of redundancy of the quantification variables. We propose here to exploit EEG source reconstruction combined to a powerful multivariate selection algorithm to improve the decoding of mental states on BCI data recorded in our lab.

2. OVERVIEW OF THE METHOD

The proposed approach is aimed at classifying in real time mental states from EEG for an asynchronous BCI. In this device, EEG signals are acquired and processed in real time on sliding time windows in 3 steps:

1. Source reconstruction: Let M be the (n,T) matrix of the measured EEG signal recorded during a time window of T samples. The corresponding amplitude of N

cortical sources, stored in the matrix $\boldsymbol{J}$ of size (N,T), is reconstructed using MNE.

2. Quantification of source activities: for each cortical source, activity is quantified by spectral power in 4 frequency bands. The quantification variables are concatenated and stored in a feature vector $\boldsymbol{x}$ of size (4N,1).
3. Classification of brain state: We focus on the case of binary classification corresponding to a subject switching between two mental tasks labeled $y = 1$ for task 1 and $y = -1$ for task 2. The ongoing mental task is then estimated by $\hat{y}$ from the feature vector using a discriminant function $f(x)$ such that $\hat{y} = \text{sign}(f(x))$.

Each step of this processing is described in the following parts.

3. SOURCE RECONSTRUCTION

Our estimate of cortical source activities rests on a forward model accounting for instantaneous EEG data formation by equivalent current dipoles whose locations and orientations are constrained to a surface tessellation of the subject's cortical mantle (surfacic model) [3, 4]. This leads to the equation:

$$\boldsymbol{M} = \boldsymbol{G}\boldsymbol{J} + \boldsymbol{\epsilon}$$

where $\boldsymbol{M}$ is the matrix of n EEG measurements on the electrode cap; $\boldsymbol{J}$ is a matrix of N elementary source amplitudes in the model; $\boldsymbol{G}$ is the gain that contains the unitary contributions of all elementary sources sampled at each electrode; $\boldsymbol{\epsilon}$ is an additive noise term. As N is very large ($\approx$ 10000) compared to n, computing an estimate $\hat{\boldsymbol{J}}$ of $\boldsymbol{J}$ from $\boldsymbol{M}$ is an ill posed inverse problem that requires additional constraints to get a unique solution. We used the Minimum Norm Estimate (MNE) [1] approach whose solution minimizes the objective

$$||\boldsymbol{M} - \boldsymbol{G}\boldsymbol{J}||^2 + \alpha||\boldsymbol{J}||^2$$

where alpha is a regularization parameter. The solution can then be expressed explicitly by

$$\hat{\boldsymbol{J}} = \boldsymbol{G}^T(\boldsymbol{G}\boldsymbol{G}^T + \alpha\boldsymbol{I}_n)^{-1}\boldsymbol{M} = \boldsymbol{W}\boldsymbol{M}$$

As the matrix $\boldsymbol{W}$ is only a function of the subjects' anatomy, this approach allows fast computation of source amplitude from EEG measurements. In our settings, according to an heuristic, alpha is set to 10% of the first eigenvalue of $\boldsymbol{G}\boldsymbol{G}^T$.

4. QUANTIFICATION

The spectral power in 4 frequency bands (8-12Hz, 15-20Hz, 20-30Hz et 30-40Hz) is computed for each cortical source using Welch's periodogram. This is done through the computation of the cross spectral matrix $\boldsymbol{\Gamma}$ for the EEG sensors in each frequency band. The spectral power in the frequency band B for source i is then computed with the equation:

$$\boldsymbol{W}_i\boldsymbol{\Gamma}(B)\boldsymbol{W}_i^T$$

Where $\boldsymbol{W}_i$ is the i-th line of $\boldsymbol{W}$. This approach requires the computations of n Fourier transforms instead of N and is thus less computationally demanding.

5. CLASSIFICATION

5.1. Basic principles

Standard classification methods can be broadly described as a two-step procedure:

- First, using a learning set $\{(\boldsymbol{x}_i, y_i)\}_{i\in I}$ consisting of feature vectors $\boldsymbol{x}_i$ and their associated class label y_i, a discriminant function $f(\boldsymbol{x})$ is optimized in order give the better prediction of the class label using the equation $\hat{y} = \text{sign}(f(\boldsymbol{x}))$: this is the training phase.

- Then the resulting discriminant function can be used to estimate class labels corresponding to new feature vectors. This is the testing phase, where the classifier accuracy is quantified by the percentage of correct estimations of the class label.

5.2. Support Vector Machine

The linear Support Vector Machine (SVM) [5] is a state of the art classification algorithm optimizing a linear discriminant function of equation $f^*(\boldsymbol{x}) = \boldsymbol{\omega}^*\boldsymbol{x} + b$ such that

$$(\boldsymbol{\omega}^*, b^*) = \arg\min_{\boldsymbol{\omega},b} \|\boldsymbol{\omega}\|^2$$

under the constraints

$$\forall i, y_i(\langle\boldsymbol{\omega}, \boldsymbol{x}_i\rangle + b) \geq 1$$

The classifier coefficients are computed by solving a dual problem in $\mathbb{R}^q$, with q the number of elements in the learning set.

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j \langle\boldsymbol{x}_i, \boldsymbol{x}_j\rangle$$

Finally the solution writes $\boldsymbol{w}^* = \sum \hat{\alpha}_i\boldsymbol{x}_i$. Usually, most of the coefficients in $\boldsymbol{w}^*$ are non-vanishing. Using this method, classifying a new EEG time window j thus requires to compute all the quantification variables of the feature vector $\boldsymbol{x}_j$. As this computation must be done in real time, reducing the number of variables exploited by the classifier is crucial.

5.3. L0 SVM variable selection

In order to select quantification variables, many methods are available. A usual technique is univariate selection with the Fisher score. Let (μ_1, μ_2) and (σ_1, σ_2) be the empirical means and standard deviations of a variable for each class. The Fisher score of this variable writes $F = \frac{(\mu_1-\mu_2)^2}{\hat{\sigma}_1^2+\hat{\sigma_2}^2}$. This quantity enables to rank the variables and select only a few of them with the highest scores to feed the classifier. A shortcoming of this univariate approach is that the selected variables may be highly redundant, whereas variables giving additional information can be discarded.

We thus used a multivariate approach through an embedded method enabling the simultaneous computation of the classifier together with the variable selection. The proposed classifier is a linear Support Vector Machine where the Euclidian norm in the objective function has been replaced by a 0 norm. The method thus finds an optimal linear discriminant function of equation $f^*(\boldsymbol{x}) = \boldsymbol{\omega}^*\boldsymbol{x} + b^*$ such that

$$(\boldsymbol{\omega}^*, b^*) = \arg\min_{\boldsymbol{\omega},b} \|\boldsymbol{\omega}\|_0$$

under the constraints

$$\forall i, y_i(\langle \boldsymbol{\omega}, \boldsymbol{x}_i\rangle + b) \geq 1$$

The 0 norm is equal to the number of non vanishing coefficients in the vector $\boldsymbol{\omega}$. This is thus equivalent to finding a linear discriminant function using a minimal number of variables. Such a method presents the interest to avoid selecting redundant variables as univariate variable selection like Fisher score could do. The exact solution of this classifier is difficult to compute, nevertheless the recursive algorithm L2AROM [6] presented below gives an approximate solution using a vector $\boldsymbol{z}$ of scaling factors associated to each variable:

1. Initialize $z = [1, 1, .., 1]^T$
2. Solve

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j y_i y_j \langle \boldsymbol{z} * \boldsymbol{x}_i, \boldsymbol{z} * \boldsymbol{x}_j\rangle$$

under the constraints $\sum \alpha_i y_i = 0$ and $\forall i, \alpha_i \geq 0$

3. Let $\hat{\boldsymbol{w}} = \sum \hat{\alpha}_i(\boldsymbol{x}_i * \boldsymbol{z})$ from the solution of 2. Iterate $\boldsymbol{z} \leftarrow \boldsymbol{z} * |\hat{\boldsymbol{w}}|$
4. Go back to 2 until convergence of $\boldsymbol{z}$. The solution is then $\boldsymbol{w}^* = \boldsymbol{z} * \hat{\boldsymbol{w}}$

This algorithm is thus equivalent to a recursive implementation of a usual linear SVM. The resulting vector contains only a few non vanishing coefficients corresponding to selected variables. This method has the advantage, compared to other embedded methods to have a computational time independent from the total number of variables since the optimization problem is stated in a dual space. The estimate of the mental tasks can then be computed in real time using the expression $\hat{y} = \text{sign}(\langle \boldsymbol{w}^*, \boldsymbol{x}\rangle + b^*)$.

6. ASYNCHRONOUS BCI EXPERIMENT

This methodology is applied offline to an asynchronous BCI experiment performed in our lab on 3 subjects. The subjects were asked to perform continuously a mental task indicated on a screen during consecutive 20s epochs. Successive tasks were separated by a resting period of 3 seconds. There were six different mental tasks, including three motor imagery tasks (grasping an object with the right hand, moving the right finger, moving the tongue) and three non-motor tasks (visuo-spatial navigation, imagined music, calculation). These tasks appeared successively in random order on the screen. For each subject, we recorded 4 to 6 sessions of 6 min each per day during 3 days. The EEG data were recorded using a 60 electrodes BrainCap. Data were amplified using a BrainAmps (Brain Products, Inc) 64 channels system sampled to 500Hz. These data are quantified on 2s time windows with an overlap of .5 s.

7. RESULTS

We performed a cross validation across the days for the three subjects and each of the 15 possible couple of tasks: classifiers were trained on one day of experiment and tested on the others, iteratively for each day, and then the average result was computed. In order to compare the classifier accuracy obtained from the source level quantification and from an electrode level quantification, the same L0SVM classifier has been applied using directly the spectral powers of each EEG sensors in the same 4 frequency bands. The performance of an univariate variable selection with the Fisher score was also investigated by preselecting 20 quantification variables at the source level (corresponding approximately to the number of variables selected using L0SVM) to feed a usual linear SVM classifier. The accuracies of the three approaches are reported for each subject on Table 1, averaged across the 15 possible discrimination tasks. The average accuracy of L0SVM at the source level reaches 75% for the best subject using the source level approach and is significantly higher than the performance obtained at the electrode level ($p<.05$) with a non parametric test realized on the 3 subjects. Moreover, the average accuracy of L0SVM is better than the accuracy obtained using Fisher score and a linear SVM. The results are encouraging as they are obtained using a prediction from one day to another using a quantification on 2s time windows. The convergence of the scaling factors during training of the classifier is exemplified on figure 1. As iterations increase, fewer variables get a high scaling factor compared to the other. Finally, only approximately 20 on the 40000 original variables are selected in 34 iterations. This illustrates the fast convergence of the algorithm in spite of the high number of initial variables.

The spatial distribution of these variables on the cortex is illustrated in Fig 2 for two subjects in interesting frequency bands for the couple of tasks associated with the best classifier

Table 1. Comparison between mean classification accuracy resulting from power features computed at the cortical sources' level compared to electrodes' level for each subject, using cross validation across days and averaged over all the possible couple of discriminant mental tasks.

	L0SVM source level	L0SVM electrode level	Fisher+SVM source level
subj 1	75.01% ± 2.0	72.76% ± 1.9	67.74% ± 2.1
subj 2	60.67% ± 1.3	61.84% ± 1.0	60.05% ± 1.2
subj 3	65.91% ± 1.2	63.12% ± 1.5	65.05% ± 1.4

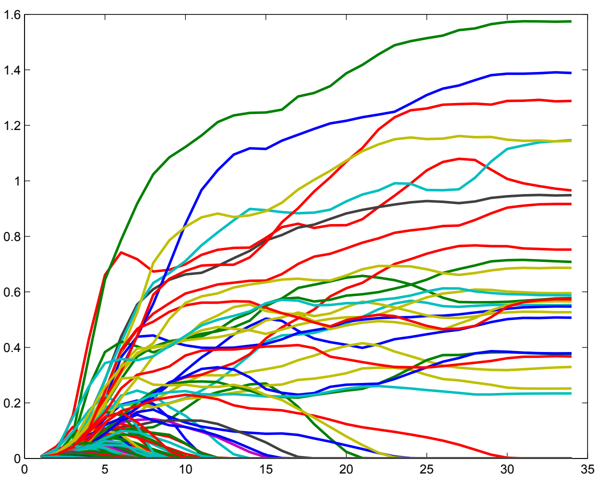
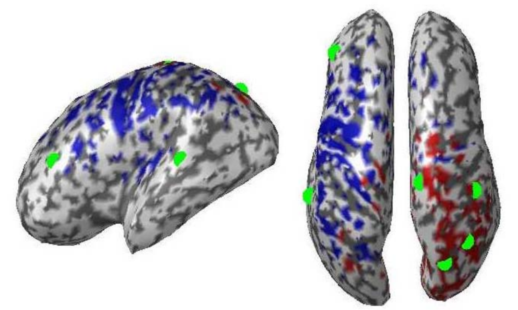


Fig. 1. Example of evolution of the scaling factors in the vector z, corresponding to each quantification variables as a function of the number of iteration of the L2AROM algorithm.

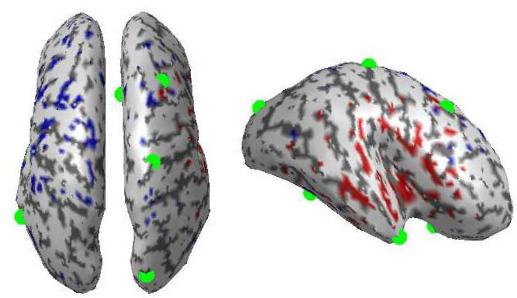
accuracy (respectively 84% for subject 1 and 75% for subject 3). The green points indicate the variables selected by the classifier whereas blue and red colors respectively represent significant increases and decreases of power in this band when switching from the first task to the second. In each case, variables are selected over the brain areas exhibiting significant increase and decrease of power, but some variables are also taken outside these regions. This possibly reflects the ability of the algorithm to select non-redundant variables avoiding then spatially correlated sources located in a same area.

8. CONCLUSION

We showed that using inverse problem can improve classification accuracies for an asynchronous BCI application. Moreover, using L0-SVM variable selection algorithm enable exploiting this approach in real time by selecting a few relevant cortical sources for an efficient quantification of the mental states.



a) Subject 1 (20-30Hz):
Tongue vs Calculation



b) Subject 3 (8-12Hz):
Visuo_Spatial vs Calculation

Fig. 2. Selected features (in green) and variables with significant t-test for the best couples of classes in 2 subjects.

9. REFERENCES

[1] S Baillet, JC Mosher, and RM Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18(6), pp. 14–30, 2001.

[2] I Guyon and A Elisseef, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[3] JJ Ermer, JC Mosher, S Baillet, and RM Leahy, "Rapidly recomputable EEG forward models for realistic head shapes.," *Phys Med Biol*, vol. 46, no. 4, pp. 1265–1281, Apr 2001.

[4] S Baillet and L Garnero, "A bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem.," *IEEE Trans Biomed Eng*, vol. 44, no. 5, pp. 374–385, 1997.

[5] V Vapnik, *Statistical Learning Theory*, Wiley-Interscience, New York, 1998.

[6] J Weston, A Elisseeff, B Schölkopf, and M Tipping, "Use of the zero norm with linear models and kernel methods," *J. Mach. Learn. Res.*, vol. 3, pp. 1439–1461, 2003.